

Pendekatan Bayesian untuk Analisis *Survival* pada Kasus Demam Berdarah *Dengue* Pasien RSUD Dr. Soetomo Surabaya

Mohammad Naufal Abdullah, Nur Iriawan dan Irhamah
Departemen Statistika, Institut Teknologi Sepuluh Nopember (ITS)
e-mail: nur_i@statistika.its.ac.id

Abstrak—Demam Berdarah *Dengue* (DBD) merupakan penyakit yang disebabkan oleh virus *Dengue* yang ditularkan melalui gigitan nyamuk *Aedes aegypti* dan *Aedes albopictus*. Indonesia menjadi negara Asia Tenggara tertinggi dengan kasus DBD. Provinsi Jawa Timur merupakan provinsi dengan jumlah kasus DBD tertinggi kedua pada 2017 dan 2018. Salah satu analisis statistika yang digunakan untuk mengetahui ketahanan hidup adalah analisis *survival*, sehingga akan dianalisis model *survival* faktor karakteristik pasien yang mempengaruhi laju kesembuhan (lama rawat inap) pasien DBD di RSUD Dr. Soetomo. Analisis Bayesian memperlakukan semua parameter yang tidak diketahui sebagai variabel *random* dan memiliki distribusi. Estimasi parameter dengan pendekatan bayesian untuk mengatasi kasus jumlah data terbatas karena mempertimbangkan distribusi *prior* (informasi sebelumnya). Model *survival* parametrik yang digunakan mengikuti pola distribusi Weibull 3 dan 2 parameter. Model terbaik dengan WAIC terkecil adalah model *survival* Weibull 2 parameter dengan faktor yang berpengaruh signifikan adalah usia pasien, pendidikan terakhir (SMA), pekerjaan (tidak bekerja), diagnosis masuk rumah sakit (II), suhu tubuh, denyut nadi, dan kadar sel darah putih.

Kata Kunci—Bayesian, Demam Berdarah *Dengue*, Lama Rawat Inap, *Survival*

I. PENDAHULUAN

DEMAM Berdarah *Dengue* (DBD) merupakan penyakit yang disebabkan oleh virus *Dengue* dari genus *Flavivirus*, famili *Flaviviridae* dan ditularkan melalui gigitan nyamuk *Aedes aegypti* dan *Aedes albopictus* [1]. Nyamuk *Aedes aegypti* menyebarkan virus *Dengue* melalui gigitannya dan virus tersebut dapat menyebabkan infeksi, mulai dari infeksi subklinis hingga infeksi simptomatik yang berakibat mulai dari demam berdarah biasa hingga demam berdarah *dengue* yang dapat meningkatkan resiko kematian [2]. Penderita demam berdarah *Dengue* akan mengalami tiga fase, yaitu fase awal, fase kritis, dan fase penyembuhan.

Nyamuk *Aedes aegypti* dan *Aedes albopictus* dapat hidup di negara tropis dan subtropis yang tersebar dari Karibia, Amerika Selatan, sebagian Afrika, Asia dan Australia serta Oseania [3]. Pada musim penghujan, virus *Dengue* berkembang cepat sehingga penyakit DBD sering menjadi kejadian luar biasa di beberapa wilayah. Indonesia menjadi negara Asia Tenggara tertinggi dengan kasus DBD [1]. Menurut data WHO, Indonesia dilaporkan sebagai negara ke-2 dengan kasus DBD terbesar diantara 30 negara wilayah endemis. Kasus demam berdarah *Dengue* (DBD) di Indonesia dari tahun 2016 sebesar 204.171 kasus turun signifikan di

tahun 2017 sebesar 68.407 kasus. Provinsi dengan jumlah kasus DBD tertinggi pada 2017 yaitu Provinsi Jawa Barat sebanyak 10.016 kasus, Jawa Timur sebanyak 7.838 kasus, dan Jawa Tengah sebanyak 7.400 kasus, serta banyak kematian kasus DBD tertinggi pada Provinsi Jawa Timur sebanyak 105 kematian [4]. Pada tahun 2018, kasus DBD di Provinsi Jawa Timur meningkat menjadi 8.449 kasus dan menduduki peringkat kedua kasus DBD tertinggi setelah Provinsi Jawa Barat. Menurut Kementerian Kesehatan hingga Januari 2019 telah terjadi 16.692 kasus DBD dengan tidak kurang dari 169 orang meninggal dunia di Indonesia. Surabaya mengalami penurunan kasus demam berdarah pada tahun 2018, akan tetapi masih terdapat beberapa kecamatan di Surabaya dengan jumlah kasus DBD yang tinggi, sehingga perlu digencarkan pencegahan kasus DBD di Surabaya, khususnya untuk wilayah Surabaya Timur.

Variabel respon pada penelitian laju atau lama kesembuhan pasien DBD adalah jarak dari awal penelitian sampai pasien DBD dinyatakan sembuh. Penelitian tersebut mendapati masalah adanya data yang tidak teramati karena keterbatasan waktu penelitian, sehingga sebagian informasi kesembuhan pasien DBD tidak diketahui secara pasti (data tersensor). Berdasarkan variabel respon laju/lama penyembuhan pasien DBD dan adanya data tersensor, maka salah satu metode analisis statistika yang digunakan pada penelitian ini adalah analisis *survival*. Untuk mengetahui faktor-faktor karakteristik pasien yang mempengaruhi laju kesembuhan pasien DBD yang diukur melalui lama rawat inap, maka perlu dilakukan penelitian secara statistik.

Berdasarkan uraian tersebut, maka perlu dilakukan penelitian untuk mengkaji waktu kesembuhan pasien penyakit DBD berdasarkan faktor karakteristik dari pasien DBD. Penelitian ini akan membahas tentang pemodelan *survival* faktor karakteristik pasien yang mempengaruhi laju kesembuhan yang diukur melalui lama rawat inap pasien demam berdarah *Dengue* di RSUD Dr. Soetomo. Estimasi parameter model *survival* spasial dilakukan dengan metode Bayesian, karena pada kondisi data sampel yang sedikit metode Bayesian dapat memberikan inferensi yang lebih baik dibandingkan metode MLE, OLS, dan EM. Penelitian ini diharapkan mampu menjadi rekomendasi dan informasi tambahan bagi RSUD Dr. Soetomo Surabaya tentang faktor-faktor karakteristik pasien yang mempengaruhi laju kesembuhan yang diukur melalui lama rawat inap pasien DBD.

II. TINJAUAN PUSTAKA

A. Pengujian Distribusi

Pengujian distribusi merupakan pengujian hipotesis yang dilakukan untuk mengetahui distribusi probabilitas suatu data. Salah satu metode pengujian distribusi data menggunakan *Kolmogorov-Smirnov* [5]. Pengujian *Kolmogorov-Smirnov* merupakan uji kesesuaian dengan menghitung selisih absolut antara fungsi distribusi kumulatif sampel yang dinotasikan $F(t)$ dan fungsi distribusi kumulatif atau *cumulative distribution function* (CDF) teoritisnya yang dinotasikan $F_0(t)$. Hipotesis uji *Kolmogorov-Smirnov*, yaitu : $H_0: F(t) = F_0(t)$,untuk semua data mengikuti distribusi yang ditentukan $H_1: F(t) \neq F_0(t)$, minimal satu data tidak mengikuti distribusi yang ditentukan Penolakan hipotesis dilakukan menggunakan nilai statistik uji *Kolmogorov-Smirnov* dapat ditulis pada (1).

$$D = \sup|F(t) - F_0(t) | \tag{1}$$

Nilai D dibandingkan dengan nilai pada tabel *Kolmogorov-Smirnov* dengan ukuran sampel n dan taraf signifikan α . Jika $D > D_{(\alpha,n)}$ atau $P\text{-value} < \alpha$ maka tolak H_0 atau data tidak mengikuti distribusi yang ditentukan.

B. Analisis Survival

Analisis *survival* merupakan metode statistik yang digunakan untuk menganalisis data dimana variabel responnya berupa waktu hingga kejadian *survival* terjadi (waktu *survival*). Kejadian *survival* yang dimaksud adalah kejadian khusus (*failure event*) yang dialami individu. Variabel respon analisis *survival* adalah lama waktu dari awal dilakukan penelitian sampai kejadian *survival* terjadi atau penelitian berakhir [6]. Terdapat tiga elemen yang perlu diperhatikan dalam menentukan waktu *survival* yaitu :

1. *Time origin* atau *starting point* (titik awal) adalah waktu dimulainya suatu penelitian.
2. *Ending event of interest* (kejadian akhir) adalah kejadian yang menjadi inti dari penelitian.
3. *Measurement scale for the passage of time* (skala ukuran untuk berlalunya waktu).

Kesulitan yang sering terjadi pada analisis *survival* adalah adanya data tersensor. Data tersensor adalah waktu *survival* (*survival time*) yang sebagian informasi ketahanan hidup individu tidak diketahui secara pasti. Klasifikasi data tersensor dibagi menjadi tiga klasifikasi, yaitu tersensor kanan, tersensor kiri, dan tersensor interval [7].

C. Fungsi Survival dan Fungsi Hazard

Model *survival* digunakan untuk menjelaskan tentang resiko terjadinya kejadian *survival* tertentu dipengaruhi oleh beberapa variabel tertetu pada selang waktu tertentu [8]. Fungsi *survival* merupakan suatu fungsi yang menyatakan peluang atau kemungkinan individu dapat bertahan lebih dari waktu tertentu t [6]. Berikut merupakan persamaan untuk fungsi *survival*.

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) \tag{2}$$

Fungsi *hazard* merupakan fungsi yang menjelaskan laju terjadinya kejadian *survival* per satuan waktu *survival*.

Fungsi *hazard* merupakan fungsi non-negatif. Fungsi *hazard* dapat dinyatakan dalam (3).

$$h(t) = \frac{f(t)}{S(t)} \tag{3}$$

Hubungan antara $H(t)$ atau fungsi kumulatif *hazard* dengan fungsi *survival* $S(t)$ adalah sebagai berikut :

$$\begin{aligned} H(t) &= -\ln S(t) \\ S(t) &= \exp(-H(t)) \end{aligned} \tag{4}$$

dengan

- t : Waktu yang diamati,
- T : Waktu *survival* seorang individu,
- $F(t)$: Fungsi distribusi kumulatif (CDF),
- $f(t)$: Fungsi kepadatan peluang (PDF),
- $S(t)$: Fungsi *survival*,
- $h(t)$: Fungsi *hazard* proporsional,
- $H(t)$: Fungsi kumulatif *hazard*.

D. Model Proportional Hazard

Model *Cox Proportional Hazard* (*Cox PH*) merupakan model semiparametrik yang dapat digunakan ketika data yang ada tidak diketahui jelas distribusinya. Hal ini dikarenakan meskipun *baseline hazard* tidak diketahui bentuk fungsionalnya, *Cox PH* tetap dapat memberikan informasi yang berguna berupa *hazard ratio* [9]. Model regresi *Cox PH* ditunjukkan pada (5).

$$h(t, \mathbf{X}) = h_0(t) \exp \left(\sum_{i=1}^p \beta_i X_i \right) \tag{5}$$

dimana,

- $h_0(t)$: *Baseline hazard* pada waktu ke- t ,
- β_i : Parameter regresi ke- i ,
- X_i : Nilai dari variabel bebas ke- i ,
- $\mathbf{X} = (X_1, X_2, \dots, X_p)$: Vektor variabel bebas.

Model terbaik diperoleh dengan melakukan estimasi koefisien variabel prediktor X_1, X_2, \dots, X_p yaitu $\beta_1, \beta_2, \dots, \beta_p$. Paramater atau koefisien regresi β pada model *Cox PH* dapat diestimasi menggunakan metode *maximum likelihood estimation* (MLE).

E. Asumsi Proporsional Hazard

Pengujian asumsi *proportional hazard* juga dapat dilakukan dengan menggunakan statistik uji *goodness of fit*. Adapun langkah-langkah untuk pengujian asumsi dengan uji *goodness of fit* adalah sebagai berikut.

1. Mendapatkan residual *schoenfeld* untuk setiap variabel prediktor dari model *cox*.
2. Membuat variabel *rank* waktu *survival* dari urutan kecil ke besar.
3. Menghitung korelasi antara residual *schoenfeld* dan *rank* waktu *survival* dengan rumus pada (6).

$$r_{RT,PR_i} = \frac{\sum_{j=1}^n (PR_{ij} - \bar{PR})(RT_j - \bar{RT})}{\sqrt{\sum_{j=1}^n (PR_{ij} - \bar{PR})^2} \sqrt{\sum_{j=1}^n (RT_j - \bar{RT})^2}} \tag{6}$$

Keterangan :

- PR_{ij} : residual *schoenfeld* untuk variabel prediktor ke- i individu yang mengalami *event* pada $t_{(j)}$.
- RT_j : *rank* waktu *survival* yang diurutkan dari kecil ke besar pada $t_{(j)}$.

Tabel 1.
Variabel Penelitian

Variabel	Nama Variabel	Skala
T	Lama rawat inap pasien	Rasio
8	Status tersensor	Nominal
X ₁	Jenis Kelamin Pasien 0 : Perempuan 1 : Laki-laki	Nominal
X ₂	Usia pasien saat dirawat	Rasio
X ₃	Pendidikan terakhir pasien 1: SMA 2 : D3/S1 3 : Lain-lain	Ordinal
X ₄	Pekerjaan pasien : 1: Tidak Bekerja 2: Pegawai Negeri 3: Pegawai Swasta 4: Lain-lain	Nominal
X ₅	Diagnosis DBD saat masuk RS : 1: I 2: II 3 : III	Ordinal
X ₆	Lama demam sebelum masuk ke RS	Rasio
X ₇	Suhu badan pasien saat masuk RS	Rasio
X ₈	Banyaknya denyut nadi per menit	Rasio
X ₉	Frekuensi pernafasan pasien	Rasio
X ₁₀	Kadar Hemoglobin pasien	Rasio
X ₁₁	Kadar Hematokrit pasien	Rasio
X ₁₂	Kadar sel darah putih pasien	Rasio
X ₁₃	Kadar trombosit pasien	Rasio

Tabel 2.
Hasil Uji *Proportional Hazard*

Variabel	rho	P-value
Jenis Kelamin (X ₁) : Perempuan	0,1333	0,4896
Usia (X ₂)	0,4151	0,0703
Pendidikan Terakhir (X ₃) : SMA	-0,2109	0,3044
Pendidikan Terakhir (X ₃) : D3/S1	-0,1998	0,2763
Pekerjaan (X ₄) : Pegawai Negeri	0,1213	0,5178
Pekerjaan (X ₄) : Pegawai Swasta	-0,0364	0,8688
Variabel	rho	P-value
Pekerjaan (X ₄) : Tidak Bekerja	NA	NAN
Lama demam sebelum masuk RS	0,33498	0,0936
Diagnosis (X ₅) : II	-0,1764	0,3400
Diagnosis (X ₅) : III	-0,0594	0,6123
Demam Sebelum di RS (X ₆)	0,2157	0,1391
Suhu (X ₇)	-0,1912	0,3339
Denyut Nadi (X ₈)	-0,1777	0,3103
Pernapasan (X ₉)	0,0346	0,7587
Kadar Hemoglobin (X ₁₀)	-0,0265	0,9009
Kadar Hematokrit (X ₁₁)	0,1607	0,3773
Kadar Trombosit (X ₁₃)	-0,3129	0,0715
GLOBAL	NA	0,5369

4. Melakukan pengujian korelasi antara residual *schoenfeld* dan *rank* waktu *survival* dengan hipotesis sebagai berikut.

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Statistik uji yang digunakan dapat dituliskan pada (7).

$$t_{hit} = \frac{r_{RT,PR_i} \sqrt{n-2}}{\sqrt{1-r_{RT,PR_i}^2}} \quad (7)$$

Tolak H₀ jika $|t_{hit}| > t_{\alpha/2, n-2}$ atau nilai P-value kurang dari taraf kesalahan (α) yang digunakan. P-value dihitung dengan rumus $2(1 - P(T \leq |t|))$. Asumsi *propotional hazard* terpenuhi jika uji korelasi antara residual *schoenfeld* dan *rank* waktu *survival* tidak signifikan atau gagal tolak H₀ [6].

Perhitungan statistik uji global *Grambsch and Thernau* yang berdasarkan residual *schoenfeld* pada regresi cox dengan hipotesis sebagai berikut :

H₀ : Asumsi *proporsional hazard* terpenuhi

H₁ : Asumsi *proporsional hazard* tidak terpenuhi

Penolakan hipotesis di atas dapat menggunakan statistik uji global variabel ke-p (G_p) dituliskan pada (8).

$$G_p = \frac{\sum_{i=1}^M \sum_{j=1}^{N_i} (\delta_{ij} h(t_{ij}) - \bar{h}) r_{s_{pij}}}{I_p \sum_{i=1}^M \sum_{j=1}^{N_i} (\delta_{ij} h(t_{ij}) - \bar{h})} \sim \chi^2_{(1)} \quad (8)$$

dengan

$$r_{s_{pij}} = \delta_{ij} \{x_{ij} - \bar{\alpha}_{pij}\}$$

$$\bar{\alpha}_{pij} = \frac{\sum_{l \in R_f} x_{pij} \exp(\beta \mathbf{X}_1)}{\sum_{l \in R_f} \exp(\beta \mathbf{X}_1)}$$

δ_{ij} menunjukkan status kejadian (tersensor (0) dan tidak tersensor (1)), I_p merupakan elemen *information matrix* variabel ke-p, $r_{s_{pij}}$ adalah nilai residual *Schoenfeld*, dan R_f merupakan himpunan waktu yang beresiko pada waktu $t_{(j)}$. Kesimpulan gagal tolak H₀, jika nilai $G_p \sim \chi^2_{(1)}$ atau jika nilai P-value ($1 - P(\chi^2 \leq G_p)$) kurang dari taraf signifikan (α) [10].

F. Hazard Ratio

Hazard Ratio merupakan suatu ukuran yang digunakan untuk mengetahui tingkat risiko perbandingan antara individu dengan kondisi variabel prediktor pada kategori sukses dengan kategori gagal [11]. *Hazard ratio* diperoleh dengan membandingkan fungsi *hazard* antar dua kategori pada satu prediktor. *Hazard ratio* dapat didefinisikan pada (9).

$$\hat{H}R(t) = \frac{\hat{h}(t, \mathbf{X}^*)}{\hat{h}(t, \mathbf{X})}$$

$$\hat{H}R(t) = \frac{\hat{h}_0(t) \exp(\sum_i^p \hat{\beta}_i X_i^*)}{\hat{h}_0(t) \exp(\sum_i^p \hat{\beta}_i X_i)} \quad (9)$$

$$\hat{H}R(t) = \exp \left[\sum_i^p \hat{\beta}_i (X_i^* - X_i) \right]$$

dimana,

$\mathbf{X}^* = [X_1^*, X_2^*, \dots, X_p^*]$ individu dengan *hazard* lebih besar

$\mathbf{X} = [X_1, X_2, \dots, X_p]$ individu dengan *hazard* lebih kecil

G. Analisis Bayesian

Pengambilan keputusan dalam statistika Bayesian berdasarkan informasi baru dari data yang diamati (sampel) dan pengetahuan sebelumnya [12]. Pendekatan statistika Bayesian mengasumsikan parameter θ adalah sebagai variabel *random* yang memiliki distribusi, disebut distribusi prior. Distribusi prior digunakan untuk mencari distribusi posterior sehingga diperoleh estimator Bayesian yang merupakan *mean* atau modus dari distribusi posterior. Apabila data observasi *survival* dinyatakan sebagai \mathbf{t}

sedangkan parameter data dinyatakan sebagai θ . Distribusi θ dengan syarat t diberikan melalui teorema Bayes dalam (10).

$$p(\theta|t) = \frac{l(t|\theta)p(\theta)}{p(t)} \quad (10)$$

dimana,

$p(\theta)$: Informasi Prior,

$l(t|\theta)$: Informasi sampel dalam *likelihood* data,

$p(\theta|t)$: Informasi posterior,

$p(t)$: Total probabilitas Bayes (konstanta penormal).

Distribusi posterior $p(\theta|t)$ merupakan perkalian *likelihood* yang berisi informasi data sampel observasi dengan prior parameter yang berisikan informasi tentang data sebelumnya, sehingga distribusi posterior dapat dituliskan seperti pada (11).

$$p(\theta|t) \propto l(t|\theta)p(\theta) \quad (11)$$

Secara garis besar terdapat empat macam distribusi prior.

- Conjugate* dan *non conjugate prior* merupakan penentuan *prior* yang didasarkan dari pola *likelihood* data.
- Informative* dan *non informative prior* merupakan penentuan *prior* yang didasarkan pada ada atau tidaknya informasi yang diperoleh dari penelitian sebelumnya mengenai pola distribusi data.
- Proper* dan *improper prior* merupakan penentuan *prior* tergantung pada pemberian pembobotan/densitas di setiap titik, apakah terdistribusi secara uniform atau tidak.
- Pseudo prior* merupakan penentuan *prior* dengan nilai yang pemberian nilainya didasarkan dan disetarakan dengan hasil elaborasi cara *frequentist*.

Pada model yang kompleks, distribusi posterior sulit untuk memperbarui (*update*) parameter dari sampel melalui distribusi posterior. Metode *Markov Chain Monte Carlo* (MCMC) dapat menyelesaikan hal tersebut. Ide dasar dari MCMC adalah membangkitkan sebuah *Markov Chain* dengan simulasi *Monte Carlo* yang diiterasikan, sehingga didapatkan distribusi posterior yang stasioner [13]. *Gibbs sampling* merupakan algoritma MCMC yang mencakup iteratif sampling dari tiap distribusi bersyarat, dimana pada model *survival* spasial Weibull 2 parameter, parameter θ dipartisi menjadi beberapa bagian yaitu $\theta = (\eta, \beta_0, \beta_2, \dots, \beta_p)$. Bentuk distribusi *full conditional* untuk masing-masing parameter yaitu

$$p(\eta|t, \mathbf{X}, \beta_0, \dots, \beta_p), p(\beta_0|t, \mathbf{X}, \eta, \dots, \beta_p), \dots,$$

$p(\beta_p|t, \mathbf{X}, \eta, \beta_0, \dots, \beta_{p-1})$ dengan t adalah data respon/*survival* dan \mathbf{X} adalah vektor dari kovariat (prediktor). Distribusi *full conditional* parameter diperoleh dengan memilah dan menghilangkan parameter lainnya pada *joint* distribusi posterior. Berikut adalah langkah-langkah dalam estimasi parameter dengan algoritma *Gibbs sampling*.

- Menentukan *initial value* dari

$$\theta^{(0)} = (\eta^{(0)}, \beta_0^{(0)}, \beta_1^{(0)}, \dots, \beta_p^{(0)}) \text{ dan } m = 0$$

- Membangkitkan tiap komponen

$$\theta^{(m+1)} = (\eta^{(m+1)}, \beta_0^{(m+1)}, \beta_1^{(m+1)}, \dots, \beta_p^{(m+1)})$$

dimana

$$\eta^{(m+1)} \text{ dari } p(\eta|t, \mathbf{X}, \beta_0^{(m)}, \dots, \beta_p^{(m)})$$

$$\beta_0^{(m+1)} \text{ dari } p(\beta_0|t, \mathbf{X}, \eta^{(m+1)}, \beta_1^{(m)}, \dots, \beta_p^{(m)})$$

$$\beta_1^{(m+1)} \text{ dari } p(\beta_1|t, \mathbf{X}, \eta^{(m+1)}, \beta_0^{(m+1)}, \dots, \beta_p^{(m)})$$

⋮

$$\beta_p^{(m+1)} \text{ dari } p(\beta_p|t, \mathbf{X}, \eta^{(m+1)}, \gamma^{(m+1)}, \dots, \beta_{p-1}^{(m+1)})$$

- Membangkitkan lebih banyak observasi dengan mengambil nilai $m = m + 1$ dan mengulangi Langkah 1 dan 2 hingga konvergen atau hingga $m = T$.
- Membuang b observasi pertama (*burn in*) untuk menghindari pengaruh *initial value*.
- Mengasumsikan $\theta^{(b+1)}, \theta^{(b+2)}, \dots, \theta^{(T)}$ (sudah membuang b observasi pertama) sebagai sampel untuk analisis posterior.
- Membentuk plot distribusi posterior.
- Mendapatkan rata-rata, median, standar deviasi, kuantil, dan *MC error* dari distribusi posterior.

H. Pemeriksaan Konvergensi

Hasil simulasi MCMC konvergen maka sampel yang dibangkitkan dari simulasi sesuai dengan distribusi posterior. Salah satu pemeriksaan konvergensi secara inferensi adalah *Gelman-Rubin*. Uji *Gelman-Rubin* merupakan pendekatan pengujian konvergensi dengan mendefinisikan beberapa rantai *Markov* dengan *initial value* yang berbeda dan membandingkan varians beberapa rantai *Markov* dengan varians masing-masing rantai *Markov*.

Jika terdapat m rantai *Markov* yang saling bebas dan masing-masing rantai *Markov* telah diambil T iterasi, maka pemeriksaan konvergensi MCMC dilakukan melalui *potential scale reduction factor* (PSRF) atau *Rhat* yang dituliskan pada (12).

$$\hat{R} = \frac{\hat{V}(\theta)}{avg} \quad (12)$$

dimana

$$\hat{V}(\theta) = \frac{T-1}{T} avg + \frac{1}{T} B \quad (13)$$

$$avg = \frac{1}{m} \sum_{i=1}^m s_i^2 \quad (14)$$

dengan

$$s_i^2 = \frac{1}{T-1} \sum_{t=1}^T (\theta_{it} - \bar{\theta}_i)^2$$

$$B = \frac{T}{m-1} \sum_{i=1}^m (\bar{\theta}_i - \bar{\bar{\theta}})^2$$

$$\bar{\bar{\theta}} = \frac{1}{m} \sum_{i=1}^m \bar{\theta}_i$$

T : Jumlah iterasi

θ_{it} : Posterior θ rantai *Markov* ke- i iterasi ke- t .

Jika nilai PSRF atau *Rhat* mendekati 1, maka setiap rantai *Markov* akan konvergen dengan distribusi yang ditargetkan. Nilai PSRF atau *Rhat* akan cenderung menurun secara ergodik mendekati 1 ketika $T \rightarrow \infty$ [14].

I. Watanabe Akaike Information Criterion (WAIC)

Watanabe Akaike Information Criterion (WAIC) atau *Widely Applicable Information Criterion* merupakan salah satu pendekatan Bayesian yang lebih lengkap untuk menduga data *out-of-sample*. Penghitungan WAIC dimulai dengan menghitung log fungsi densitas *pointwise* posterior (pposterior) dan menambahkan koreksi jumlah parameter efektif untuk menyesuaikan *overfitting*. Rumus untuk menghitung WAIC dapat dituliskan pada (15).

$$WAIC = -2elppd_{WAIC} \quad (15)$$

dengan

$$elppd_{WAIC} = lppd - pWAIC$$

$$lppd = \sum_{i=1}^n \log \left(\frac{1}{T} \sum_{m=1}^T p(x_i | \theta^{(m)}) \right)$$

$$pWAIC = \sum_{i=1}^n var_{post}(\log p(x_i | \theta))$$

T : Jumlah iterasi.

WAIC memiliki kelayakan yang diinginkan rata-rata di atas distribusi posterior dibandingkan dalam satu estimasi titik, jika dibandingkan dengan AIC dan DIC. Ukuran kebaikan model dengan WAIC sangat relevan digunakan dalam prediksi, karena WAIC mengevaluasi prediksi yang sebenarnya untuk data baru dalam Bayesian [15].

J. Demam Berdarah Dengue (DBD)

Demam Berdarah *Dengue* (DBD) merupakan yang disebabkan oleh virus *Dengue* dari genus *Flavivirus*, famili *Flaviviridae* yang ditularkan ke manusia melalui gigitan nyamuk *Aedes aegypti* yang terinfeksi oleh virus *Dengue*. Tiga faktor utama penularan virus *Dengue*, yaitu faktor manusia, virus, dan vektor perantara. Kasus DBD pertama kali masuk ke Indonesia pada tahun 1968 di Surabaya [1]. Penyakit DBD biasanya ditandai dengan demam selama 2-7 hari, lemah/lesu, nyeri pada ulu hati, nyeri pada sendi, dan pendarahan pada kulit (bintik-bintik). DBD merupakan penyakit yang sering terjadi pada negara yang memiliki iklim tropis dan sub-tropis yang tersebar mulai Amerika Selatan, Afrika hingga Asia Tenggara. DBD memiliki tiga fase, yaitu fase demam, fase kritis, dan fase penyembuhan. Fase demam merupakan fase ketika penderita DBD mengalami demam tinggi 1-3 hari. Demam penderita DBD pada fase kritis cenderung turun, namun hal tersebut berakibat pada penurunan trombosit secara drastis yang apabila salah dalam menangani dapat menyebabkan gangguan peredaran darah. Trombosit penderita DBD kembali naik pada fase penyembuhan.

III. METODOLOGI PENELITIAN

A. Sumber Data

Pada penelitian ini sumber data yang akan digunakan merupakan data sekunder yang diperoleh dari data rekam medis pasien rawat inap DBD di RSUD Dr. Soetomo Surabaya. Waktu penelitian adalah 1 Januari 2016 sampai dengan 31 Desember 2019.

B. Variabel Penelitian

Variabel-variabel yang digunakan dalam penelitian ini terdapat pada Tabel 1.

C. Langkah Analisis

Langkah-langkah analisis dalam penelitian ini dituliskan sebagai berikut :

1. Melakukan pengujian asumsi pemodelan *proportional hazard*.
2. Melakukan pengujian distribusi lama rawat inap pasien DBD.
3. Menentukan model *survival* lama rawat inap pasien DBD di Surabaya Timur dengan faktor-faktor yang mempengaruhi menggunakan simulasi MCMC *Gibbs Sampling*. Langkah-langkahnya dapat dijelaskan sebagai berikut; (a) Menentukan distribusi *prior* yang digunakan berdasarkan informasi dari data; (b) Menentukan *joint posterior* model *survival*; (c) Menentukan *full conditional* untuk setiap parameter; (d) Melakukan *initial value* untuk parameter awal; (e) Melakukan simulasi MCMC dengan algoritma *Gibbs Sampling* hingga hasil simulasi konvergen.
4. Membentuk fungsi *hazard* pasien DBD berdasarkan posterior *summaries* yang diperoleh.
5. Interpretasi model *survival* sehingga diperoleh faktor-faktor yang berpengaruh terhadap lama rawat inap pasien DBD.
6. Menarik kesimpulan dan saran.

IV. ANALISIS DAN PEMBAHASAN

A. Karakteristik Data

Analisis deskriptif dilakukan untuk mengetahui karakteristik pasien DBD yang menjalani rawat inap di RSUD Dr. Soetomo Surabaya. Pasien DBD di RSUD Dr. Soetomo asal Surabaya Timur rata-rata waktu rawat inap sampai dinyatakan sembuh selama 4 hari, dengan lama rawat inap paling cepat adalah 1 hari, dan rawat inap paling lama adalah 7 hari. Pasien tergolong usia dewasa dan mengalami gejala demam sebelum ke rumah sakit selama 3-7 hari. Mayoritas pasien DBD memiliki kadar Trombosit yang cukup rendah. Terdapat pasien dengan suhu tubuh yang cukup tinggi, pasien dengan frekuensi denyut nadi di atas angka normal, dan pasien yang memiliki kadar Hematokrit di atas 46%.

Pada penelitian kali ini, pasien DBD yang menjalani rawat inap di RSUD Dr. Soetomo didominasi oleh pasien laki-laki, pasien dengan tingkat pendidikan SMA, pasien yang bekerja sebagai pegawai swasta, dan pasien yang masuk diagnosis I. Pasien dengan tingkat pendidikan D3/S1, semua berjenis kelamin laki-laki. Pasien yang bekerja sebagai pegawai swasta didominasi pasien laki-laki, sedangkan pasien yang bekerja sebagai PNS dan tidak bekerja didominasi pasien perempuan. Pasien laki-laki, tidak bekerja, dan bekerja sebagai pegawai swasta lebih banyak masuk kelompok diagnosis I. Satu pasien perempuan, pendidikan terakhir SMA, dan tidak bekerja masuk diagnosis III.

Tabel 3.
Uji Distribusi Waktu *Survival*

Distribusi	Ranking	P-value	Keputusan
Weibull 3 Parameter	14	0,7803	Gagal tolak H_0
Weibull 2 Parameter	18	0,6715	Gagal tolak H_0

Tabel 4.
Nilai WAIC

Model	WAIC
Model <i>Survival</i> Weibull 3 Parameter	72,5
Model <i>survival</i> Weibull 2 Parameter	71,8

Tabel 5.
Hazard Ratio Model *Survival* Terbaik

Variabel	Estimasi	Pembanding	Hazard Ratio	1/Hazard Ratio
Usia Pasien	0,0142	10 tahun	1,1524	0,8677
Pendidikan (SMA)	-0,6774	1 Tingkat	0,5080	1,9686
Pekerjaan (Tidak Bekerja)	-0,6348	1 Tingkat	0,5301	1,8866
Diagnosis (II)	-1,0669	1 Tingkat	0,3441	2,9064
Suhu Tubuh	-0,2138	5 Derajat	0,3434	2,9123
Denyut Nadi	0,0324	10 Kali	1,3821	0,7235
Kadar Sel	0,0000	0,0000	0,0000	0,0000

B. Asumsi Proportional Hazard

Uji asumsi *proportional hazard* pada penelitian ini menggunakan uji *goodness of fit Grambsch and Thernau*. Tabel 2 merupakan hasil uji *goodness of fit* asumsi *proportional hazard*.

Hasil uji asumsi *proportional hazard* menunjukkan untuk seluruh variabel prediktor yang digunakan dalam penelitian ini memiliki *P-value* lebih dari taraf signifikan sebesar 0,05, sehingga gagal tolak H_0 . Uji Global asumsi *proportional hazard* untuk keseluruhan model secara simultan memiliki *P-value* lebih dari taraf signifikan sebesar 0,05. Hal tersebut menunjukkan bahwa secara statistik asumsi *proportional hazard* telah terpenuhi. Nilai rho atau korelasi pada variabel pekerjaan (tidak bekerja) adalah NA, hal ini terjadi karena adanya dependensi pada variabel tingkat pendidikan (X_3) dan pekerjaan (X_4), serta tercatat pada data bahwa pasien yang tidak bekerja semua berpendidikan terakhir SMA, sehingga parameter untuk pekerjaan (X_4) kategori tidak bekerja tidak diestimasi atau dihilangkan (NA). Pada analisis model *survival* akan dilakukan dengan pendekatan Bayesian karena dapat mengatasi dependensi antar variabel prediktor, sehingga variabel pekerjaan (X_4) khususnya kategori tidak bekerja dapat dilakukan estimasi dan tidak dihilangkan.

C. Pendugaan Distribusi Lama Rawat Inap

Pendugaan distribusi waktu lama rawat inap pasien DBD Surabaya Timur di RSUD Dr. Soetomo dilakukan dengan uji *Kolmogorov-Smirnov* dan asumsi awal sebagai hipotesis nol (H_0) adalah data lama rawat inap pasien DBD mengikuti pola distribusi Weibull 2 atau 3 parameter. Pemilihan dua distribusi tersebut karena distribusi Weibull sering digunakan analisis *survival* dan model *survival* parametrik dari Weibull mudah diinterpretasikan.

Hasil uji kesesuaian distribusi pada Tabel 3 menunjukkan bahwa waktu lama rawat inap pasien DBD Surabaya Timur di RSUD Dr. Soetomo mengikuti pola distribusi Weibull 2 dan 3 parameter.

D. Joint Posterior dan Distribusi Prior

Pada penelitian ini digunakan dua distribusi untuk model *survival* parametrik lama inap pasien DBD Surabaya Timur di RSUD Dr. Soetomo yaitu distribusi Weibull 2 dan 3 parameter. Fungsi kepadatan peluang distribusi Weibull 3 parameter dapat dituliskan pada (16).

$$f(t|\eta, \mu, \gamma) = \frac{\eta}{\mu^\eta} (t - \gamma)^{\eta-1} \exp\left(-\left(\frac{t-\gamma}{\mu}\right)^\eta\right) \quad (16)$$

Persamaan (17) merupakan fungsi *hazard* untuk distribusi Weibull 3 parameter.

$$h(t_{ij}, \mathbf{X}_{ij}) = \eta(t_{ij} - \gamma)^{\eta-1} \exp \beta_0 + \beta_1 X_{ij1} + \dots + \beta_p X_{ijp} \quad (17)$$

sebelum melakukan estimasi terhadap parameter η , γ dan untuk β_i , maka terlebih dahulu menentukan distribusi *prior* masing-masing parameter sebagai berikut

$$\eta \sim \text{Gamma}(r, s)$$

$$\gamma \sim \text{Normal}(r, s)$$

$$\beta_i \sim \text{Normal}(v, w)$$

Joint distribusi posterior parameter model *survival* spasial Weibull 3 parameter dapat dituliskan pada (18).

$$p(\beta_0, \dots, \beta_p, \eta, \gamma | t_{ij}, \mathbf{X}_{ij}, \delta_{ij}) = \quad (18)$$

$$\prod_{i=1}^s \prod_{j=1}^{n_i} f_{ij}(t_{ij})^{\delta_{ij}} S_{ij}(t_{ij})^{1-\delta_{ij}} p(\beta_0) \dots p(\beta_p) p(\eta) p(\gamma)$$

Persamaan (19) merupakan fungsi kepadatan peluang distribusi Weibull 2 parameter.

$$f(t|\eta, \mu) = \frac{\eta}{\mu^\eta} t^{\eta-1} \exp\left(-\left(\frac{t}{\mu}\right)^\eta\right) \quad (19)$$

dengan fungsi *hazard* pada (20).

$$h(t_{ij}, \mathbf{X}_{ij}) = \eta t_{ij}^{\eta-1} \exp(\beta_0 + \beta_1 X_{ij1} + \dots + \beta_p X_{ijp}) \quad (20)$$

Distribusi prior parameter η dan β_i sama seperti pada model Weibull 3 parameter. Penentuan distribusi *prior* tersebut dilakukan berdasarkan gabungan antara *prior* informatif dan *pseudo prior*. *Joint* distribusi posterior parameter model *survival* spasial Weibull 2 parameter dapat dituliskan pada (21).

$$p(\beta_0, \dots, \beta_p, \eta | t_{ij}, \mathbf{X}_{ij}, \delta_{ij}) \quad (21)$$

$$= \prod_{i=1}^s \prod_{j=1}^{n_i} f_{ij}(t_{ij})^{\delta_{ij}} S_{ij}(t_{ij})^{1-\delta_{ij}} p(\beta_0) \dots p(\beta_p) p(\eta)$$

E. Faktor-faktor yang Signifikan Mempengaruhi Lama Rawat Inap Pasien DBD dengan Model Weibull 3 Parameter

Faktor-faktor yang signifikan mempengaruhi lama rawat inap pasien DBD di RSUD Dr. Soetomo adalah jika nilai *credible interval* 2,5% hingga 97,5% tidak memuat 0. Faktor karakteristik pasien yang berpengaruh terhadap lama rawat inap pasien DBD adalah usia pasien, pendidikan terakhir pasien (SMA), pekerjaan pasien (Tidak Bekerja), diagnosis

DBD saat masuk ke rumah sakit (II), suhu tubuh, denyut nadi per menit, dan kadar sel darah putih.

F. Faktor-faktor yang Signifikan Mempengaruhi Lama Rawat Inap Pasien DBD dengan Model Weibull 2 Parameter

Faktor-faktor karakteristik pasien yang mempengaruhi lama rawat inap pasien DBD di RSUD Dr. Soetomo secara signifikan pada model Weibull 2 parameter adalah usia pasien, pendidikan terakhir pasien (SMA), pekerjaan pasien (Tidak Bekerja), diagnosis DBD saat masuk ke rumah sakit (II), suhu tubuh, denyut nadi per menit, dan kadar sel darah putih.

G. Perbandingan Model Survival

Model *survival* terbaik adalah model yang memiliki nilai WAIC yang lebih kecil. Tabel 4 merupakan nilai WAIC dari masing-masing model *survival*. Berdasarkan Tabel 4 diperoleh informasi bahwa model *survival* Weibull 2 parameter memiliki nilai WAIC terkecil, yaitu 71,8. Sehingga dapat disimpulkan bahwa model *survival* terbaik adalah model *survival* Weibull 2 parameter.

H. Interpretasi Model Survival Lama Rawat Inap Pasien DBD

Model *survival* Weibull 2 parameter dapat dituliskan pada (22).

$$h(t_i) = 3,4423_i^{2,4423} \exp(0,8214 + 0,0142X_2 - 0,677X_{3(SMA)} - 0,6348X_{4(Tidak Bekerja)} - 1,0669X_{5(II)} - 0,2138X_7 + 0,0324X_8 + 0,1270 X_{12}) \quad (22)$$

Parameter $\eta > 1$, sehingga peningkatan laju kesembuhan (laju rawat inap) pasien DBD seiring dengan peningkatan waktu lama rawat inap pasien DBD. *Hazard Ratio* model *survival* Weibull 2 parameter untuk faktor yang berpengaruh signifikan disajikan pada Tabel 5.

Tabel 5 menunjukkan informasi bahwa pasien dengan usia 10 tahun lebih tua memiliki lama rawat inap 1,1524 kali lebih cepat. Lama rawat inap pasien dengan tingkat pendidikan SMA 1,9686 lebih lama dari pasien dengan tingkat pendidikan lainnya (D3/S1 keatas). Lama rawat inap pasien yang tidak bekerja juga lebih lama 1,8866 kali dibandingkan pasien yang bekerja lainnya, sehingga pasien yang tidak bekerja rawat inapnya cenderung lama. Peningkatan suhu tubuh pasien sebesar 5°C rawat inapnya lebih lama 2,9123 kali dan peningkatan jumlah denyut nadi 10 kali per menit, lama rawat inapnya lebih cepat 1,3821 kali. Pasien dengan penambahan 3,5 ribu sel darah putih akan memiliki lama rawat inap 1,5596 kali lebih cepat. Pasien DBD yang masuk rumah sakit dengan diagnosis II memiliki rawat inap lebih lama 2,9064 kali dari diagnosis I, maka pasien dengan diagnosis masuk II rawat inapnya relatif lama. Namun secara medis dan WHO, semakin tinggi level diagnosis maka akan semakin lama waktu kesembuhannya. Pada penelitian kali ini, hanya terdapat satu pasien DBD asal Surabaya Timur yang masuk diagnosis III, sehingga pola lama rawat inap pasien yang masuk diagnosis III tidak terlihat dan menyebabkan pasien yang masuk diagnosis II memiliki waktu lama rawat inap yang lebih lama dibandingkan diagnosis III. Berdasarkan kondisi tersebut, dapat diberikan saran pada penelitian serupa untuk menghilangkan observasi

yang dapat menjadikan kontradiksi pada pengambilan keputusan atau membatasi pada ruang lingkup penelitian tertentu.

V. KESIMPULAN

Kesimpulan yang diperoleh dari hasil analisis yang telah dilakukan adalah Pasien DBD di RSUD Dr. Soetomo asal Surabaya Timur rata-rata waktu rawat inap sampai dinyatakan sembuh selama 4 hari, dengan lama rawat inap paling cepat adalah 1 hari, dan rawat inap paling lama adalah 7 hari. Pasien tergolong usia dewasa dan mengalami gejala demam sebelum ke rumah sakit selama 3-7 hari. pasien DBD yang menjalani rawat inap di RSUD Dr. Soetomo didominasi oleh pasien laki-laki, pasien dengan tingkat pendidikan SMA, pasien yang bekerja sebagai pegawai swasta, dan pasien yang masuk diagnosis I.

Asumsi *proportional hazard* telah terpenuhi dan lama rawat inap pasien DBD mengikuti pola distribusi Weibull 2 dan 3 parameter. Model *survival* parametrik terbaik adalah model *survival* Weibull 2 parameter. Faktor karakteristik pasien yang berpengaruh terhadap lama rawat inap pasien DBD adalah usia pasien, pendidikan terakhir pasien (SMA), pekerjaan pasien (Tidak Bekerja), diagnosis DB ketika masuk ke rumah sakit (II), suhu tubuh, denyut nadi per menit, dan kadar sel darah putih.

Berdasarkan kesimpulan yang diperoleh, dapat diberikan saran untuk penelitian berikutnya yaitu perlu dipertimbangkan lagi faktor-faktor yang diduga mempengaruhi lama rawat inap (laju kesembuhan) pasien DBD, perlu mempertimbangkan dan menambahkan efek spasial untuk penelitian selanjutnya, pemodelan *survival* parametrik menggunakan distribusi yang rankingnya di atas distribusi Weibull 3 dan 2 parameter.

Rekomendasi atau informasi tambahan yang dapat diberikan kepada RSUD Dr. Soetomo adalah mengutamakan atau segera memberikan perawatan lebih pada pasien DBD yang masuk diagnosis II (level diagnosis yang lebih tinggi), pasien cenderung memiliki suhu tubuh tinggi, denyut nadi rendah, dan kadar sel darah putih rendah agar laju sembuh pasien meningkat (lama rawat inap cenderung cepat). Pencerdasan khusus pada pasien DBD yang pendidikan terakhir SMA dan tidak bekerja mengenai prosedur penanganan DBD secara medis dan mandiri, serta bekerjasama dengan pemerintah untuk memberikan pencerdasan dan pengetahuan tentang pencegahan DBD pada masyarakat, khususnya yang pendidikan terakhir SMA dan tidak bekerja.

DAFTAR PUSTAKA

- [1] Kemenkes RI, "Jendela Epidemiologi Demam Berdarah Dengue,," Kementerian Kesehatan RI, Jakarta, 2010.
- [2] Q. Jing and M. Wang, "Dengue epidemiology," *Glob. Heal. J.*, vol. 3, no. 2, 2019, doi: 10.1016/j.glohj.2019.06.002.
- [3] S. Leta, T. J. Beyene, E. M. De Clercq, K. Amenu, M. U. G. Kraemer, and C. W. Revie, "Global risk mapping for major diseases transmitted by *Aedes aegypti* and *Aedes albopictus*," *Int. J. Infect. Dis.*, vol. 67, 2018, doi: 10.1016/j.ijid.2017.11.026.
- [4] Pusat Data dan Informasi Kemenkes RI, "Situasi Penyakit Demam Berdarah Tahun 2017," Kementerian Kesehatan RI, Jakarta, 2018.
- [5] A. M. Law and W. D. Kelton, *Simulation Modelling Analysis*, 3rd ed. New York: McGraw-Hill, 2000.
- [6] D. G. Kleinbaum and M. Klein, *Statistics for Biology and Health*

- Survival Analysis: A Self-Learning Text*, 3rd ed. New York: Springer, 2012.
- [7] D. Collet, *Modelling Survival Data in Medical Research*, 3rd ed. London: Chapman and Hill, 2015.
- [8] D. Darmofal, "Bayesian spatial survival models for political event processes," *Am. J. Pol. Sci.*, vol. 53, no. 1, 2009, doi: 10.1111/j.1540-5907.2008.00368.x.
- [9] I. N. Aini, "Extended Cox Model untuk Time Independent Covariat yang Tidak Memenuhi Asumsi Proportional Hazard pada Model Cox Proportional Hazard.," Universitas Indonesia, 2011.
- [10] P. M. Grambsch and T. M. Therneau, "Proportional hazards test and diagnostics based on weighted residuals," *Biometrika*, vol. 81, no. 3, pp. 515–526, 1994.
- [11] D. W. Hosmer, S. Lemeshow, and S. May, *Applied Survival Analysis: Regression Modelling of Time to Event Data*, 2nd ed. Canada: John Willey & Sons, Inc., 2008.
- [12] K. S. Chin, D. W. Tang, J. B. Yang, S. Y. Wong, and H. Wang, "Assessing new product development project risk by Bayesian network with a systematic probability generation methodology," *Expert Syst. Appl.*, vol. 36, no. 6, 2009, doi: 10.1016/j.eswa.2009.02.019.
- [13] D. Sorensen and D. Gianola, *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. New York: Springer, 2002.
- [14] S. P. Brooks and G. O. Roberts, "Convergence assessment techniques for Markov chain Monte Carlo," *Stat. Comput.*, vol. 8, no. 4, pp. 319–335, 1998, doi: 10.1023/A:1008820505350.
- [15] A. Gelman, J. Hwang, and A. Vehtari, "Understanding predictive information criteria for Bayesian models," *Stat. Comput.*, vol. 24, no. 6, pp. 997–1016, 2014, doi: 10.1007/s11222-013-9416-2.